

# UC Berkeley

## UC Berkeley Previously Published Works

### Title

Phylogenetic classification supports a Northeastern Amazonian Proto-Tupí-Guaraní Homeland

### Permalink

<https://escholarship.org/uc/item/35f5k43f>

### Authors

O'Hagan, Zachary  
Chousou-Polydouri, Natalia  
Michael, Lev

### Publication Date

2022-12-22

### DOI

10.20396/liames.v19i0.8655791

Peer reviewed

## Phylogenetic classification supports a Northeastern Amazonian Proto-Tupí-Guaraní homeland

Zachary O'Hagan

University of California, Berkeley, US

<https://orcid.org/0000-0002-2720-2070>

Natalia Chousou-Polydouri

University of Zurich, Zurich, CH

<https://orcid.org/0000-0002-5693-975X>

Lev Michael

University of California, Berkeley, US

<https://orcid.org/0000-0002-9612-5996>

**ABSTRACT:** The question of where Proto-Tupí-Guaraní (PTG) was spoken has been a point of considerable debate. Both northeastern and southwestern Amazonian homelands having been proposed, with evidence from both archaeology and linguistic classification playing key roles in this debate. In this paper we demonstrate that the application of linguistic migration theory to a recent phylogenetic classification of the Tupí-Guaraní family lends strong support to a northeastern Amazonian homeland.

**KEYWORDS:** Tupí-Guaraní; Historical Linguistics; Phylogenetics; Proto-language Homeland.

**RESUMEN:** La cuestión de dónde fue hablado el proto-tupí-guaraní (PTG) ha sido un punto de debate considerable. Se han propuesto puntos de origen en el noreste y en el suroeste amazónicos, sustentados mayormente a través de evidencia arqueológica y de clasificaciones lingüísticas. En este artículo demostramos que la aplicación de la teoría de migración lingüística a una reciente clasificación filogenética de la familia tupí-guaraní favorece la hipótesis de un punto de origen amazónico nororiental.

**PALABRAS CLAVES:** Tupí-Guaraní; Lingüística histórica; Filogenética; Territorio de protolengua.

### 1. Introduction

The Tupí-Guaraní (TG) family is striking for its great geographical extent, and the study of movements of TG peoples over historical time scales has correspondingly been an important theme in TG anthropology, ethnohistory, and archaeology (see Noelli 1998, 2008 for an overview). It is also clear that the TG expansion has significantly shaped the linguistic, cultural and social history of lowland South America (see, e.g., Haynie et

al. 2014; Michael 2014), making the question of where Proto-Tupí-Guaraní (PTG), the ancestor of all modern TG languages, was spoken, and how the languages diversified and radiated across South America an important question for diverse fields engaged with the indigenous peoples of the continent.

In this paper we apply linguistic migration theory (LMT) to: 1) the geographical distribution of modern TG languages; and 2) the most fine-grained and empirically well supported classification of the family, Michael et al.'s (2015) phylogenetic TG classification, to locate the PTG homeland and to clarify key aspects of the dispersal of TG languages. We show that this method indicates a northeastern Amazonian homeland for PTG, supporting the claims of archaeologists such as Lathrap (1970) and Brochado (1984), but contradicting those of archaeologists such as Iriarte et al. (2017), and linguists such as Rodrigues (2000). We also identify a particular subgroup of the TG family as having been especially spatially dynamic, spreading TG languages both west along the upper Amazon, and south along the Atlantic coast and then eastwards into the Paraná River basin and beyond.

Specifically, we propose that the PTG homeland was located on the lower Xingu River, and that several of the major high-level subgroups resulted from splits that took place either within the Xingu basin, or relatively nearby, in adjacent river basins, and near the mouth of the Amazon River. We find that one major subgroup, which we label Diasporic, expanded across much of the continent, spreading up the Amazon (Omagua and Kukama), along the Atlantic Coast (Tupinambá), and southwards (the Southern group, which includes the Guaranian subgroup).

In the remainder of this paper we describe the data and methodology employed (§2); present the results, including both the inferred PTG homeland and observations about the geographical radiation of PTG's daughter languages (§3); contextualize these results with respect to previous scholarship on these questions (§4); explicitly compare the plausibility of northeastern and southwestern PTG homeland hypotheses in light of the application of LMT to the Michael et al. (2015) classification; and then conclude (§6).

We close this introduction with a caveat: note that in this paper we talk about the homeland for a *proto-language*, not a historical *people* or *population*, and the dispersal<sup>1</sup> of *languages*, not *peoples*. This choice is deliberate: while it is of course true that languages only exist and move through space as a consequence of being learned and used by speakers, it is potentially problematic to assume that the movement of languages corresponds directly to the movement of peoples. In particular, processes of language shift can result in changes in the spatial distributions of languages without significant population movements (Dyen 1956; Nichols 1997a, *inter alia*). The ultimate question of how the diversification and dispersal of Tupí-Guaraní languages corresponds to the movement of cultural practices and populations through space and time is a larger project that will require synthesis of research in archaeology, ethnography, ethnohistory, human

---

<sup>1</sup> In the context of discussing the movements of Arawakan peoples, Heckenberger (2002) argues for using the term 'diaspora', rather than 'dispersal', since the former emphasizes the agency of the peoples involved. We find compelling Heckenberger's argument as applied to the diversification and movement of peoples, but feel that the term 'dispersal' is a better choice for the spatio-temporal trajectories of languages, about which we feel we should be more cautious in attributing agency.

genetics and linguistics. This paper is a contribution to the linguistic component of this necessarily collaborative endeavor.

## 2. Methodology

The methodology we employ to infer the PTG homeland and the trajectories of the dispersal of TG languages is linguistic migration theory (LMT), which was first developed by Edward Sapir (1916) and later formalized by Isidore Dyen (1956) (see also Diebold 1960; Nichols 1997a). LMT is a qualitative, abductive, parsimony-based method for determining the homelands of proto-languages based on the spatial distribution of modern languages and the genealogical relationships among them. The reversal of this abductive process yields hypothetical trajectories by which the proto-language's daughters spread, resulting in the modern distribution of languages. This methodology has been applied to identifying the homelands and dispersal trajectories of the Austronesian (Blust 2013; Pawley and Ross 1993), Indo-European (Nichols 1997b), and Athabascan (Dumond 1969) language families, among others.

The key idea behind LMT is that, all other things being equal, the best explanation for the current distribution of genealogical units that are coordinate in a family tree is the one that requires the least independent movement from the region posited to be where their immediate ancestor was spoken. For example, suppose we have four genealogical groups -- A, B, C, and D -- that are coordinate in a family tree, where A, B, and C are spoken in close proximity to each other, and D is spoken in a region distant from the one in which A, B, and C are spoken. The LMT leads to the hypothesis that the proto-language was spoken in the region where A, B, and C are spoken, and that D moved to its more distant location, as opposed to the hypothesis that the proto-language was spoken in the region where D is spoken, and that all of the other groups moved independently from that region to cluster together far from the proto-language homeland. This is what yields the common principle that that locus of highest genealogical diversity in a language family should be taken as the homeland for that family. This is exemplified by the famous case of Austronesian, where the presence of multiple top-level Austronesian subgroups on the island of Taiwan, and only one of them, Malayo-Polynesian, being distributed elsewhere, leads to the conclusion that Taiwan was the Proto-Austronesian homeland (Pawley and Ross 1993, *inter alia*).

The abductive processes of LMT lead not only to inferences about the regions where proto-languages were spoken; the reversal of them also leads to inferred trajectories by which daughter languages came to occupy their modern locations. For example, in the above hypothetical example, we infer that D migrated from the proto-language homeland to its modern location. And by inferring the locations of mid-level proto-languages, it is possible to more precisely delineate the plausible trajectories associated with the dispersal of successive daughters of the proto-language.

In both the inference of homelands and trajectories of dispersal, cultural and geographical factors may inform the abductive process. For example, if all the societies associated with languages of a particular subgroup are strongly oriented around aquatic transport and the exploitation of aquatic resources, as is the case of the Omagua-Kukama-

Tupinambá subgroup we discuss below, dispersal trajectories along waterways would be, all other things being equal, more plausible than overland trajectories.

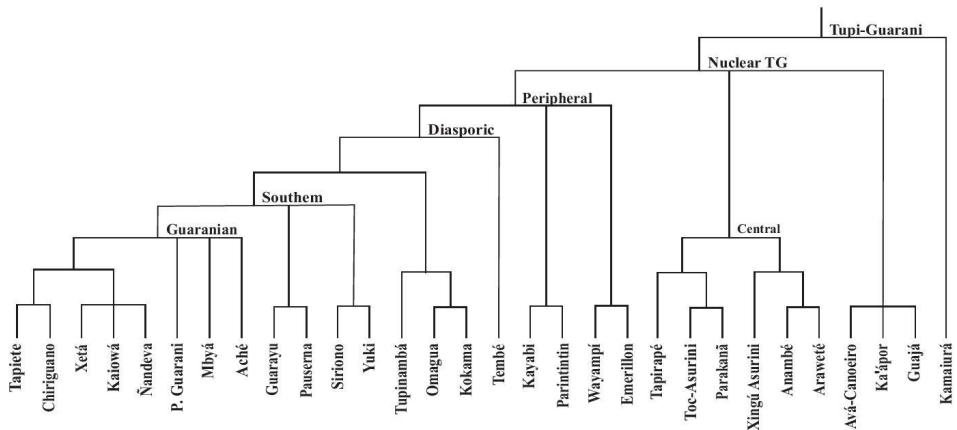
We conclude this section by observing that the reliability of the results we obtain from the application of the LMT to Michael et al.'s (2015) classification of the TG family is contingent on two factors: 1) the accuracy of classification of the family; and 2) the degree to which the dispersal of the TG languages is parsimonious, in the sense assumed by the LMT. On the first point, Michael et al.'s (2015) classification is the most detailed and empirically best-substantiated classification currently available, but we should expect empirically and analytically improved classifications to emerge in coming years, as more scholars turn to the important question of the internal classification of the TG family. Whether new classifications will result in any significant changes to the model of the dispersion of the TG languages presented in this paper remains to be seen. On the second point, LMT, as a methodology, yields what one could call the 'simplest' model for the dispersal of languages within any language family. However, the agency exhibited by human societies should make us cautious about assuming that their spatial dynamics in contexts of language diversifications always follow the parsimony assumptions of the LMT. While it strikes us as highly unlikely that the spatial dispersal of TG societies diverged so markedly from the assumptions of the LMT as to affect the major results presented in this paper, e.g., the northeastern Amazonian PTG homeland, and the proposed dispersal trajectories of its major subgroups, it would not surprise us if data from affine disciplines – especially archaeology – will lead to reevaluation of details of the model.

## **2. Classification and distribution of languages**

As described in §2, the empirical bases of the analyses carried out in this paper are: 1) the classification of TG languages; and 2) their spatial distribution. The classification we employ is Michael et al.'s (2015) conservative<sup>2</sup> classification of TG resulting from the Bayesian phylogenetic analysis of lexical data corresponding to a 543-item concept list, which is reproduced in Fig. 1. Several of the major subgroups that we discuss below are labeled in this classification.

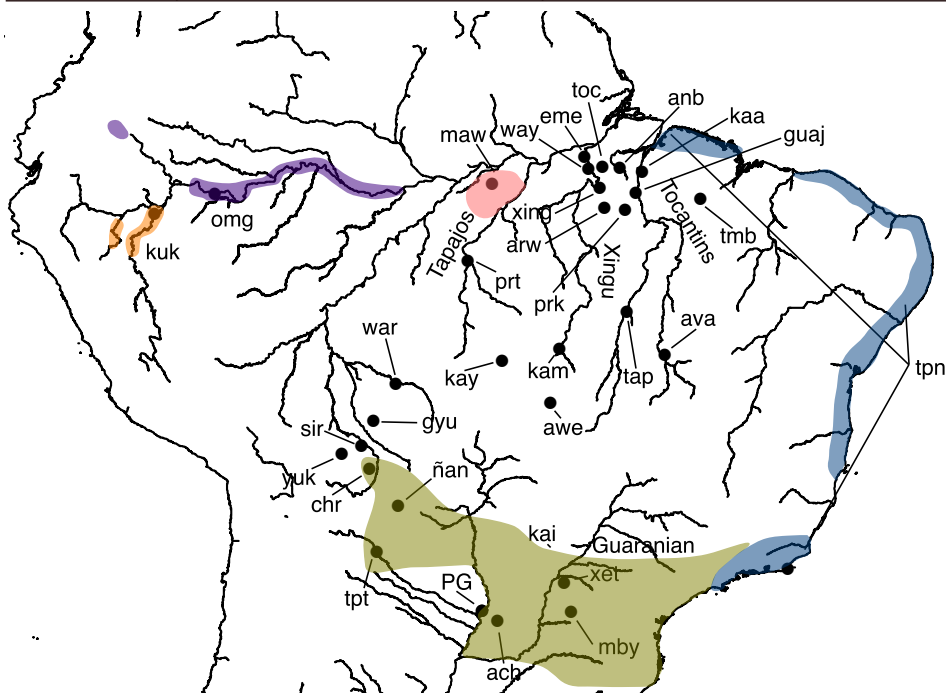
---

<sup>2</sup> In their conservative classification, Michael et al. (2015) disregard any subgroup with a posterior probability of less than 0.80.



**Figure 1:** Michael et al.'s (2015) conservative classification of the TG family

The spatial distribution of TG languages that we employ in this analysis is given in Fig. 2. Significantly, these are not the modern distributions of the languages in question (some of which, like Tupinambá, are in fact extinct), but rather their ‘time of contact’ (ToC) locations, which constitute their earliest known locations. We employ ToC locations instead of modern locations for two main reasons: 1) the locations of some languages have changed significantly since Europeans’ arrivals in South America; and 2) the geographical extent occupied by some languages has shrunk considerably. In both cases, ToC locations and distributions are better bases for inferring homelands and trajectories than modern ones, which reflect histories of genocide, displacement, and resistance that obscure the relationships of some languages to the homelands in which mid-level proto-languages were spoken. The most significant differences between ToC and modern distributions include: 1) Emerillon and Wayampí, now spoken in French Guyana and northern Amapá, respectively, which were spoken on the lower Xingu in the early colonial period (Grenand 1982); Guajá and Ka’ápor, now both spoken in the state of Maranhão, which were probably spoken on the lower Tocantins not long before the arrival of Europeans (Balée 1994); 3) Tupinambá, now extinct, which was spoken along much of the Atlantic coast of Brazil south of the mouth of the Amazon River; and Omagua, which is now on the verge of extinction, but which was spoken along a significant extent of the upper Amazon (Michael 2014; Michael and O’Hagan 2016).



**Figure 2:** Time-of-contact distribution of TG languages analyzed by Michael et al. (2015)

All languages are of course spoken in territories with spatial extent, but for purposes of representational convenience we have for the most part opted to represent the location of languages as points. The only cases where we have not done so involve languages whose spatial extent is so great that using points to indicate their location would be a gross misrepresentation. One such case is that of Tupinambá, whose spatial extension along a significant fraction of the Brazilian coast cannot be adequately represented with a single point. Correspondingly, the spatial distribution is represented by a number of blue polygons (see Fig. 2).

The one case where we depart somewhat from the use of points and polygons as outlined above is that of the Guaranian languages, where we combine point representations with a larger polygon. The points represent the approximate location of the modern languages, while the polygon represents the approximate ToC distribution of this set of closely related languages. Our motivation for combining these two representational schemes for this group of languages is the lack of clarity regarding the distinctness of all the Guaranian varieties, and their location, at ToC. We believe that this hybrid representation allows to more informatively capture both the modern distinctness of the varieties represented by points and remain suitably agnostic about their distinctness at ToC.

### 3. Analysis

We now turn to the Linguistic Migration Theory (LMT) analysis, based on the classification given in Fig. 1, and the spatial distributions given in Fig. 2. In §3.1 we use

the distribution of the major high-level subgroups to determine the location of the Proto-Tupí-Guaraní (PTG) homeland. In §3.2 we turn to clarifying the dispersal trajectories of the successive daughter languages of PTG, which requires us to identify the homelands of a number of mid-level proto-languages which were not necessary to identify the PTG homeland.

### 3.1. The Proto-Tupí-Guaraní homeland

Michael et al.'s (2015) classification of TG exhibits two major coordinate branches at the root, a single-member branch consisting solely of Kamayurá, and Nuclear-TG (NTG), that is, the remainder of the family. We begin by inferring the homeland of Proto-NTG (PNTG), since it has a number of branches that facilitate LMT inferences, and then employ the location of Kamayurá, as well as the location of the most closely-related non-TG Tupian languages, Awetí and Mawé, to infer the PTG homeland.

NTG exhibits three coordinate branches: 1) the small subgroup consisting of Avá-Canoeiro, Ka'apor, and Guajá, which call the Tocantins subgroup; 2) the larger Central subgroup; and 3) the remainder of NTG, which call the Peripheral subgroup, due to its members marking the periphery of the vast TG expansion. We proceed with our inference of the PNTG homeland by inferring the Proto-Central and Proto-Tocantins homelands, and then bring in the distribution of the Peripheral subgroups to infer the Proto-NTG homeland.

Beginning with the inference of the Proto-Central homeland, we observe from Fig. 3a that all the languages of the Central subgroup cluster in the lower Xingu River (or immediately adjacent to it), with the exception of Tapirapé, which is found further south, on the middle Araguaia River. From this we infer the lower Xingu River basin as the Proto-Central homeland. Two of the Tocantins languages, Guajá and Ka'apor are found immediately to the east, in the lower Tocantins River basin, with Avá found further upriver, on the upper Tocantins. Since these three languages are coordinate, we infer the lower Tocantins River basin to be the Proto-Tocantins homeland. The plausibility of our inferences for the Proto-Central and Proto-Tocantins homelands is enhanced by the fact that the homelands for these two coordinate branches are in immediately adjacent river basins, yielding a straightforward geographical basis for the split.



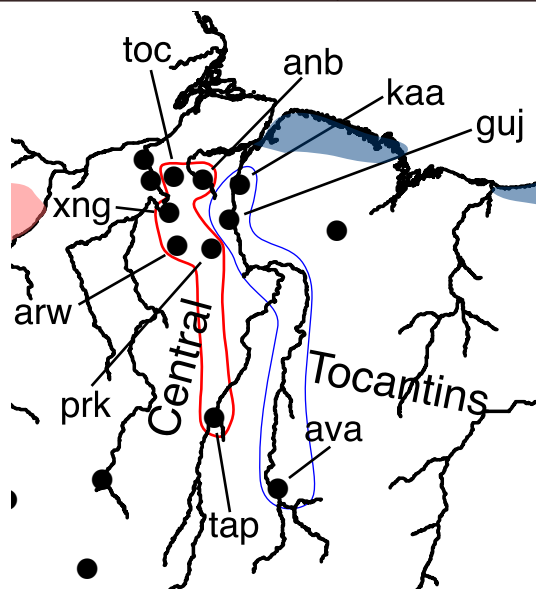


Figure 3a: Distribution of Central and Tocantins subgroups

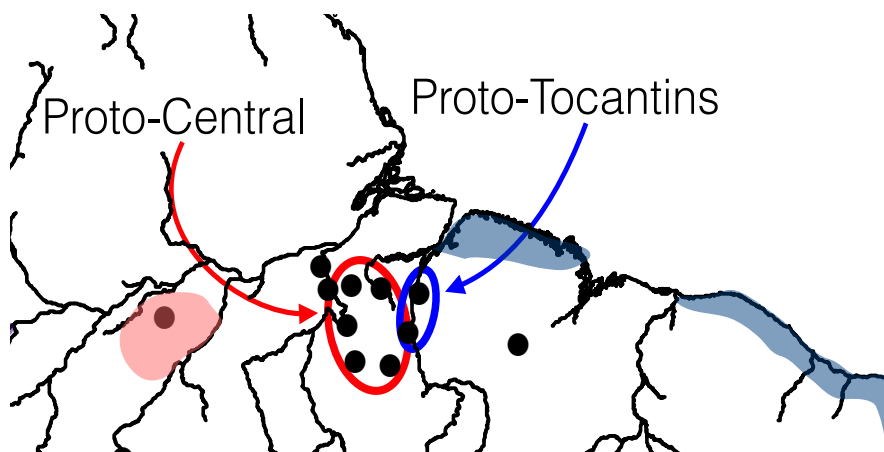


Figure 3b: Proto-Central and Proto-Tocantins homelands

We now turn to the Peripheral subgroup, observing that Peripheral consists of three major subgroups: 1) the Kayabi-Parintintin subgroup; 2) the Emerillon-Wayampí subgroup; and 3) Diasporic, which encompasses the remaining languages. We will briefly defer the question of the Proto-Peripheral homeland to return to our current main concern: The Proto-NTG homeland. The critical observation here is that one of the Peripheral subgroups, the Emerillon-Wayampí subgroup, was, at time of contact, located on the lower Xingu, adjacent to the Proto-Central homeland. This makes the lower Xingu-

Tocantins area the region of greatest genealogical diversity for the NTG subgroup,<sup>3</sup> since it is the homeland of both the proto-Central and proto-Tocantins subgroups, and was the ToC location for one of the three first-order NTG subgroups. From this, we conclude that the lower Xingu-Tocantins area is the Proto-NTG homeland, as depicted in Fig. 4.

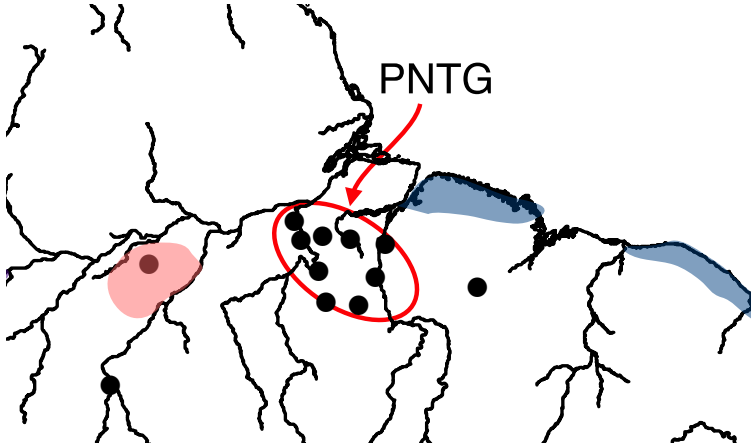


Figure 4: The Proto-Nuclear Tupí-Guaraní Homeland

Having inferred the Proto-NTG homeland, we now address the question of the PTG homeland, and will then return to the question of the Proto-Peripheral homeland and the diversification and dispersal of the Peripheral languages in §3.2.

Recall that TG consists of two coordinate subgroups, NTG and the single member Kamayurá subgroup. Significantly, Kamayurá is located on the middle Xingu River, while the inferred NTG homeland encompasses the lower Xingu and Tocantins River basins. This distribution suggests that the Proto-Tupí-Guaraní (PTG) homeland was located in the Xingu River basin. The inference is reinforced by the fact that the first sister to TG within the Tupian stock, Awetí (Galúcio et al. 2015, Rodrigues and Cabral 2012), is likewise found in the Xingu River basin. This strongly suggests a scenario where the Proto-Awetí-Tupí-Guaraní (PATG) homeland was located in the Xingu River basin, and where the split between Pre-Awetí and PTG involved the languages separating within the Xingu River basin, as was the case for the split between Pre-Kamayurá and Proto-NTG.

Having inferred that the PTG homeland lies in the Xingu River basin, we now consider whether we can further narrow its location. Here we argue that the geographical location of the Mawé and Mundurucuic branches of Tupian make the lower Xingu, rather than, say, the middle or upper Xingu, the most plausible homeland for PTG.

<sup>3</sup> A stronger hypothesis would be to posit the lower Xingu River basin alone as the Proto-NTG homeland. Given the proximity of the Xingu and Tocantins Rivers in their lower reaches, however, and the fact that the preference for the Xingu homeland rests on only one of the Peripheral subgroups being found there, we prefer to be conservative and posit the wider region encompassing both the Proto-Central and Proto-Tocantins subgroups as the Proto-NTG homeland.

We first step back and observe that there is little doubt that, as Rodrigues (2012, *inter alia*) has argued, Rondônia is the Proto-Tupian homeland, given that it is the locus of the family's genealogical diversity. Furthermore, it is clear that Mawé is a sister to the Awetí-TG subgroup, forming the Maweti-TG subgroup (Corrêa da Silva 2010; Meira and Drude 2015; Rodrigues and Dietrich 1997). Likewise, the Mundurukuic branch of Tupian is classified as a sister to the MATG subgroup both in expert<sup>4</sup> classifications (e.g., Rodrigues and Cabral 2012: 496) and in distance-based phylogenetic classifications of the family (Galúcio et al. 2015).

Turning now to the spatial distributional facts, we observe that these languages are either spoken on the lower Tapajós, or in adjacent areas that are easily reached from the lower Tapajós. The two Mundurukuic languages, Mundurukú and Kuruáya, are located on the lower Tapajós and lower Xingu, respectively, where the western tributaries of the lower Xingu meet the eastern tributaries of the lower Tapajós,<sup>5</sup> forming an easily transited corridor between the two river basins. Mawé, on the other hand, is located in the region between the lower Tapajós and lower Madeira Rivers (Nimuendajú 1948b), which is drained by rivers that flow into the Amazon proper. Significantly, the headwaters of these rivers abut the lower Tapajós, and the mouths of these rivers are a relatively small distance upriver, on the Amazon proper, from the mouth of the Tapajós. Mawé territory is thus likewise connected to the lower Tapajós – in fact by two easily transited corridors.

These spatial distributional facts suggest that the following is the simplest dispersal scenario, depicted in Fig. 5: Proto-Mundurukuic-Maweti-TG (PMMATG) moved from the vicinity of the Proto-Tupian homeland in Rondônia towards northeastern Amazonia, sooner or later moving into the Tapajós basin, where PMMATG split into Proto-Mundurukuic and PMATG. Proto-Mundurukuic subsequently split, with Pre-Mundurukú remaining mainly in the Tapajós basin, and Pre-Kuruáya moving a small distance east into the lower Xingu basin via the tributaries of these two rivers, which virtually meet. PMATG then split into Pre-Mawé and Proto-Awetí-TG (PATG). This split mostly likely occurred on the lower Tapajós, or on the Amazon River proper, near the mouth of the Tapajós. The reason for inferring this is that: 1) the inferred homeland for the node above PMATG, i.e., PMMATG, is the lower Tapajós; and 2) the ToC (and modern) location of one of the branches resulting from the PMATG split, i.e., Pre-Mawé, is a small distance east of the lower Tapajós. The proposed lower Tapajós location for the split between Pre-Mawé and PATG split has the virtue of requiring the least movement from the inferred location of the split of the immediately higher node (i.e., the Proto-Mundurukuic-PMATG split) and minimizes movement to the ToC location of one of the coordinate branches resulting from the split of PMATG itself, i.e., the ToC location of Mawé, the descendant of Pre-Mawé.

---

<sup>4</sup> 'Expert' classifications are based on linguists' deep knowledge of a language family, but not on explicitly presented subgrouping criteria and evidence; see Michael and Chousou-Polydouri (to appear) for further discussion.

<sup>5</sup> At the time that Europeans encountered the Mundurukú, they were in the midst of expanding their territory further northwards along the Tapajós River (Horton 1948). (Nimuendajú suggests that their original territory was located on the Rio das Tropas (personal communication cited in Horton *ibid.*: 273), an eastern tributary of the Tapajós towards the southern limit of their modern territory.) Nimuendajú (1948a) places the Kuruáya at ToC farther to the north than their late 19th-century location on the Curuá River, which is a tributary of the Iriri River, one of the largest tributaries of the lower Xingu.

Regardless of whether the Pre-Mawé-PATG split occurred on the lower Tapajós or nearby on the Amazon proper, the modern locations of Mawé and the ATG subgroup suggest that Pre-Mawé moved a small distance to the west to the tributaries of the Amazon between the Tapajós and Madeira Rivers, while PATG moved a small distance to the east, to the Xingu River basin. Crucially, the shortest movement, on this scenario, would have been to the lower Xingu, not the middle or upper Xingu, supporting the conclusion that the PATG homeland, and thus the PTG homeland, was located on the lower Xingu.

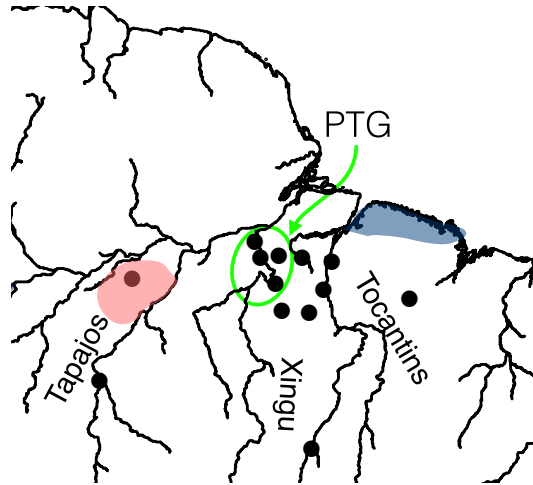


Figure 5: Proto-Tupí-Guarani Homeland

### 3.2. Dispersal of the TG languages

In the previous section we argued that the application of LMT to Michael et al.'s (2015) TG classification, combined with information about the immediate sisters to TG, leads to the inference that the PTG homeland was located in the lower Xingu River basin. Now we reverse the LMT inference process to understand the dispersal of the TG languages across the continent.

The first steps of the dispersal process follow directly from reversing the LMT reasoning process that led us to locate the PTG homeland on the lower Xingu. First, PTG is located on the lower Xingu (and additionally spreads to the lower Tocantins) and splits into Proto-Nuclear TG (PNTG), which remains on the lower Xingu, and Pre-Kamayurá, which eventually migrates upriver to its ToC location, as depicted in Fig. 6.

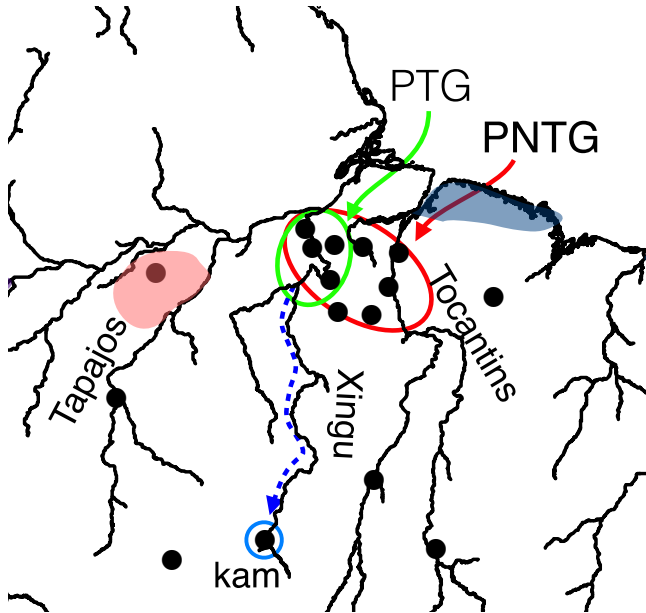


Figure 6: Dispersals associated with split of PTG into PNTG and Pre-Kamayurá

Second, PNTG then undergoes a three-way split into Proto-Peripheral, Proto-Central, and Proto-Tocantins, with Proto-Central located on the lower Xingu, Proto-Tocantins on the lower Tocantins River basin, and Proto-Peripheral in the region east of the Tocantins River basin as depicted in Fig. 7.

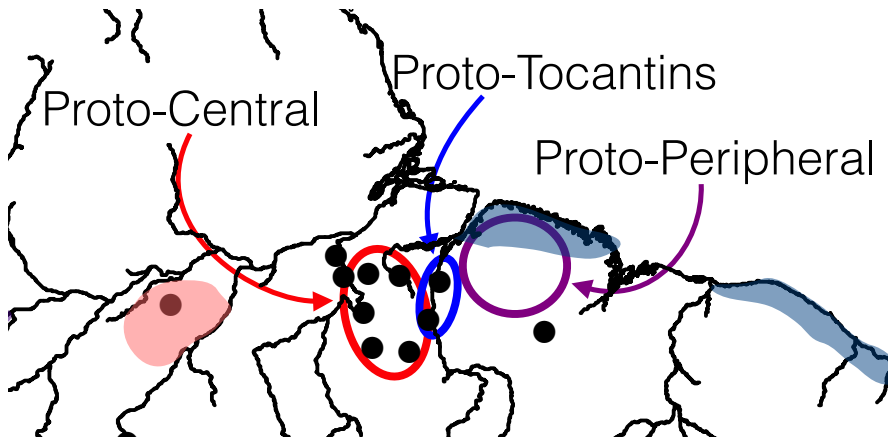


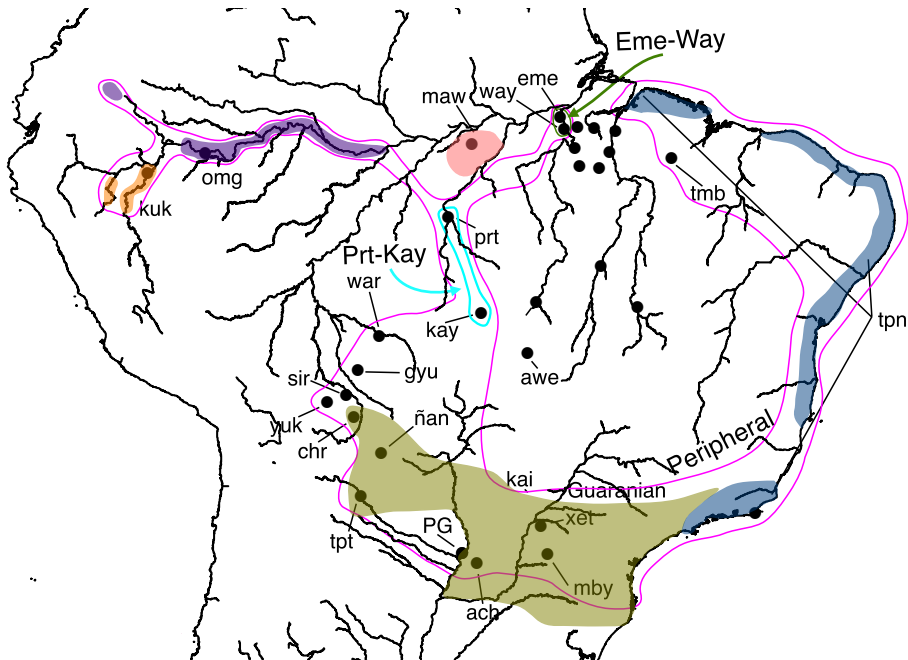
Figure 7: Split of PNTG into Proto-Peripheral, Proto-Central, and Proto-Tocantins

We now need to address the issue of the Proto-Peripheral homeland, since the distribution of the Peripheral languages is quite extensive. As evident from Fig. 8, the Peripheral subgroup encompasses most of the TG languages, and is so named not because it is unimportant -- far from it -- but because its members are distributed around much

of the periphery of TG territory in South America. Tupinambá, for example occupied much of the Atlantic coast of Brazil at ToC, the easternmost extent of the TG family, while Omagua and Kukama occupied significant stretches of the upper Amazon River basin, marking the northwestern limit of the TG expansion. Similarly, the languages of the Southern subgroup group mark the southern limit of the family.

In order to address the question of the Proto-Peripheral homeland, we will need to carry out LMT inferences on Peripheral and its various subgroups, with a special focus on the Diasporic subgroup, which is responsible for much of the geographic extent of Peripheral, and within Diasporic, the Southern subgroup, which represents a major geographic extension of the languages of the Diasporic subgroup.

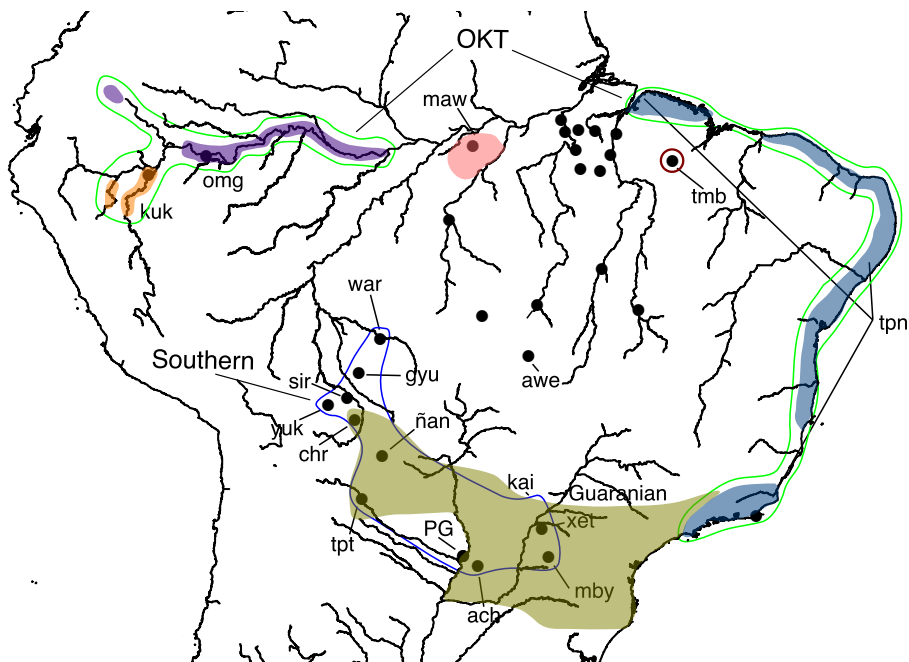
To begin, we observe that Peripheral experienced a three-way split into: 1) the Parintintin-Kayabí subgroup, whose ToC location is centered on the confluence of the Arinos and Jurueña Rivers (Nimuendajú 1924, 1948c); 2) the Emerillon-Wayampi subgroup, which was located on the lower Xingu at ToC; and 3) the large Diasporic subgroup, where ToC distribution of these subgroups is given in Fig. 8. Given the vast distribution of the Diasporic subgroup languages, inferring the Proto-Peripheral homeland requires that we first determine the Proto-Diasporic homeland, the question to which we now turn.



**Figure 8:** First-order subgroups of the Peripheral subgroup (Peripheral bounded by pink line; all languages not in the Eme-Way or Prt-Kay subgroups are members of the Diasporic subgroup)

We begin by observing that Diasporic itself splits into three groups: 1) Tembé, 2) the Omagua-Kukama-Tupinambá (OKT) subgroup; and 3) the large Southern subgroup, whose ToC distributions are given in Fig. 9. We will address the question of the Proto-

Diasporic homeland by first determining the proto-Southern homeland and then integrating the ToC locations of Temb  and the OKT group.



**Figure 9:** First-order subgroups of the Diasporic subgroup; note OKT (outlined in green) is discontinuous, and Temb  (tmb) is a single-member subgroup

Southern consists of three coordinate subgroups, the Yuki-Sirion  subgroup, the War zu-Guaray  subgroup, and the large Guaranian subgroup, as depicted in Fig. 10. The Guaranian subgroup is centered on the Paran  and Paraguay River basins, with a number of varieties located at the periphery of these basins, or outside it to the west, such as Chiriguano and Tapiete. In contrast, the War zu-Guaray  subgroup is located in the Guapor  River basin, and the Yuki-Sirion  subgroup located in the Mamor  River basin, both relatively far to the west of the Paran -Paraguay River basin. A mechanical application of LMT would suggest a proto-Southern homeland somewhere in an area spanning the upper Mamor  and Guapor  River basins, as this area encompasses both the Yuki-Sirion , War zu-Guaray , and one member of the Guaranian subgroup, making it the region of highest genealogical diversity for the Southern subgroup.



Figure 10: First-order subgroups of the Southern subgroup

There are some reasons to be cautious about this conclusion, however. First, we know from ethnohistorical sources that the arrival of Chiriguano, the westernmost Guaranian language, in the Andean foothills region dates to only the 14th or 15th century as a result of an east-to-west expansion (Santos-Granero 2009). Similarly, Tapiete has been argued to emerge as a consequence of ‘Guaranization’ of non-TG-peoples from the 16<sup>th</sup> century on (Combès 2008). Both these observations suggest that the expansion of the Guaranian subgroups was not from the Andean foothills region towards the Paraná-Paraguay River basin, but the reverse, and that this in fact took place relatively recently, with the Paraná-Paraguay River basin being the proto-Guaranian homeland. Note that this is consistent with the fact that the greatest number of coordinate branches of the Guaranian subgroup are found in the Paraná basin, suggesting that Proto-Guaranian was spoken there, and that as the Guaranian subgroup diversified, some of its members spread eastwards towards the Andean foothills. This insight undercuts the observation above regarding the genealogical diversity found in the upper Mamoré-Guaporé region.

The second reason to be cautious about positing that the upper Mamoré-Guaporé region as the proto-Southern homeland comes from Nichols’ (1997a) observation that the effect of multiple successive spreads from a common center within a given spread zone is increased diversity at the edge of the spread zone, which is precisely the circumstances we find with the languages of the Southern group. Coupled to our conclusion regarding the nature of the Guaranian expansion from the Paraná-Paraguay River basin, this suggests that it was in fact the Paraná-Paraguay River basin that was the proto-Southern homeland, with two earlier spreads bringing the ancestors of the Yuki-Sirionó and Warázu-Guarayú subgroups into the Mamoré and Guaporé basins, with any other daughter languages pertaining to these subgroups, were there any, having been absorbed by the subsequent Guaranian spread.



For these reasons, we infer that the Proto-Southern homeland was located in the Paraná-Paraguay River basin, and that the presence to the west of this basin of the non-Guaranian Southern subgroups and Guaranian languages such as Chiriguano and Tapiete was due to successive westward spreads.

Having identified the Proto-Southern homeland, we can now return to the question of the Proto-Diasporic homeland. Diasporic, it will be recalled, consists of three first-order subgroups: 1) Tembé; 2) the Omagua-Kukama-Tupinambá (OKT) subgroup; and 3) the Southern subgroup. Like the Southern subgroup, the OKT subgroup, although consisting of only three languages, extended over a large region: Omagua and Kukama in the upper Amazon, and Tupinambá along the Atlantic coast. Given the central role of aquatic resources and transportation for the groups speaking these languages, the languages clearly spread via major waterways: The Amazon proper for Proto-Omagua-Kukama (POK; Michael, 2014; O'Hagan 2011, O'Hagan 2019a,b) and the Atlantic littoral in the case of the Pre-Tupinambá. It is known from archaeological evidence that the culture associated with POK arrived in the upper Amazon in approximately 1100CE, after a steady progression from points downriver (Lathrap 1970), indicating a migration from the lower Amazon region. Tupinambá, on the other hand, was distributed from territory along southern banks of the Amazon, near the mouth of the river, to large portions of the Atlantic coast south of the Amazon. This suggests a POKT homeland in the vicinity of the lower Amazon region, with POK having migrated upriver and Pre-Tupinambá having expanded southwards along the coast. This conclusion is reinforced by the fact that one of the Diasporic sisters to the OKT subgroup, Tembé, is found in this area, specifically (at ToC), in the region that is now the state of Maranhão.

From these observations, we infer that the Proto-Diasporic homeland was located near the mouth of the Amazon, and south of the river, since two of the first-order daughters of Proto-Diasporic, POKT and Pre-Tembé, were spoken in this region, with only the third first-order daughter, Proto-Southern, spoken outside it. Given the ToC distributions of the relevant groups, LMT does not allow us to clarify whether the Proto-Diasporic homeland was located towards the west, near the mouths of the Xingu or Tocantins rivers, or further to the east, near the Atlantic coast. Regardless, it follows from the location of the Proto-Diasporic homeland in this region that the first order split involved proto-Southern moving far to the south, while POKT and Pre-Tembé remained near the mouth of the Amazon. POKT subsequently split into POK and Pre-Tupinambá, with the former moving far up the Amazon River, and the latter extending south along the Atlantic coast, but crucially also remaining in the area near the mouth of the Amazon River.

An important question that remains unanswered by the Diasporic dispersal scenario just sketched out concerns the route by which Proto-Southern reached the Paraná-Paraguay basin. As observed by Urban (1996), an Amazonian PTG homeland is compatible with multiple plausible routes by which the Southern languages could have reached their ToC locations, including a southward coastal route, followed by an inland western route, and southward routes along a number of Amazonian tributaries, such as the Tocantins or Xingu rivers, followed by an overland route from the headwaters of these rivers to the headwaters of the Paraná-Paraguay basin. Resolution of this question remains an important priority for interdisciplinary research on the history of TG peoples, cultures, and languages.

Having identified the Proto-Diasporic homeland, we can now address the location of the Proto-Peripheral homeland. Recall that Peripheral consists of three first-order subgroups, the Kayabí-Parintintin subgroup, the Emerillon-Wayampí subgroup, and Diasporic. Given the proximity of the Proto-Diasporic homeland and the ToC locations of Emerillon and Wayampí, LMT leads us to conclude that the Proto-Peripheral homeland was somewhere in the region spanned by these two first order groups, i.e., somewhere between the lower Xingu, in the west, and the Atlantic coast near the mouth of the Amazon, in the east. We argue that within this relatively large region, the eastern portion of this region, i.e., the territory east of the mouth of the Tocantins, extending to the Atlantic coast, is the most likely region for the Proto-Peripheral homeland.

Our inferring this location for the Proto-Peripheral homeland is based on the geographic distribution of the first order subgroups of the next higher node, i.e., Proto-Nuclear-TG (PNTG), and a model for the diversification of NTG. Recall that PNTG has three first-order daughters: Proto-Central, with a homeland on the lower Xingu, Proto-Tocantins, with a homeland on the lower reaches of the Tocantins, and Peripheral itself. Given that the Xingu and Tocantins River basins are each occupied by a daughter of PNTG, we argue it is less likely that Proto-Peripheral shared one of these river basins, than it having its own distinct territory. That territory, by this reasoning, and the above delimitation of the possible area in which Proto-Peripheral was spoken, would have to be the territory to the east of the Tocantins River.

Note that if we posit that the Proto-Peripheral homeland was located in the territory east of the Tocantins basin, an attractive model for the diversification of NTG follows: in brief, once PTG split into Pre-Kamayurá and PNTG, PNTG expanded rapidly eastward from the lower Xingu River basin towards the Atlantic coast, occupying a large territory from the Xingu River basin in the west, to the territory near the Atlantic coast, in the east. The three daughters of PNTG -- Proto-Central, Proto-Tocantins, and Proto-Peripheral -- then simply correspond to descendants of each segment of the PNTG-speaking population that were isolated by the major river system boundaries, i.e., the populations in the Xingu basin, Tocantins basin, and the area east of the Tocantins basin, respectively.

With the inference of the Proto-Peripheral homeland in hand, we now summarize the diversification and dispersal processes described above:

1) Proto-Mundurukuic-Mawetí-Guaraní (PMMATG) arrives in the lower Tapajós River basin and splits into Proto-Mundurukuic and Proto-Mawetí-Guaraní (PMATG)

i. Proto-Mundurukuic splits into Pre-Mundurukú, which remains in the Tapajós River basin, and Pre-Kuruáya, which moves a small distance east to the lower Xingu River basin.

2) PMATG splits in Pre-Mawé and Proto-Awetí-Guaraní (PATG) in the vicinity of the lower Tapajós River

i. Pre-Mawé moves a small distance west, to the tributaries of the Amazon between the mouths of the Tapajós and Madeira Rivers

ii. PATG moves a small distance east to the lower Xingu River

3) PATG splits into Pre-Awetí and PTG on the lower Xingu

- i. Pre-Awetí moves up the Xingu to its ToC location
- ii. PTG remains on the lower Xingu

4) PTG splits into Pre-Kamayurá and Proto-Nuclear TG (PNTG)

- i. Pre-Kamayurá moves up the Xingu to its ToC location
- ii. PNTG remains downriver, expanding to cover the region extending from the lower Xingu River basin in the west, to the territory near the mouth of the Amazon, and east of the Tocantins River basin, in the east

5) PNTG splits into Proto-Central, Proto-Tocantins, and Proto-Peripheral

- i) Proto-Central diversified in the lower Xingu and lower Tocantins River basins region
- ii) Proto-Tocantins diversified in the Tocantins River basin
- iii) Proto-Peripheral diversified in the region east of the Tocantins River basin

6) Proto-Peripheral splits into Proto-Kayabí-Parintintin (PKP), Proto-Emerillon-Wayampí (PEW), and Proto-Diasporic

- i) PKP and PEW both migrate westward, PKP reaching the southwest reaching the region of the confluence of the Arinos and Juruena Rivers, while PEW only travels as far and Xingu River basin
- ii) Proto-Diasporic diversifies in the region east of the Tocantins River basin

7) Proto-Diasporic splits into Pre-Tembé, Proto-Omagua-Kukama-Tupinambá (POKT), and Proto-Southern.

- i) Pre-Tembé essentially does not move, being already close to its ToC location
- ii) POKT splits into Pre-Proto-Omagua-Kukama (Pre-POK) and Pre-Tupinambá, with Pre-POK making its way up the Amazon to eventually arrive near the ToC location of Omagua, and Pre-Tupinambá beginning a steady expansion southward along the Atlantic coast.
- iii) Proto-Southern moves to the Paraguay-Paraná River basin, where it diversifies into Proto-Guaraní, Proto-Siriono-Yuki, and Proto-Warázu-Guarayú. All three diversify, with the latter two spreading westwards first, and the Guaranian languages following.

#### **4. Previous language classification-based theories regarding the PTG homeland**

The question of the geographical origin of the Tupí-Guaraní peoples has been an important one in South American archaeology and anthropology, marked by considerable debate, up to the present (see e.g., Almeida and Neves 2015). Numerous scholars have addressed this question, but as Noelli (1998) usefully summarizes in his cogent overview of the relevant scholarship, the varied proposals that have been made mainly fall into one of two groups: those that posit a southwestern origin, centered on the Paraná River basin, and those that posit a northeastern Amazonia origin. The evidence for these proposals comes from a variety of sources, including modern material culture, archaeological remains, and in some cases, language, where the spatial dynamics of languages is taken to be a reliable

indicator of the movements of Tupí-Guaraní peoples (a position about which, we remind the reader, we are much more cautious).

In this section we briefly review the two works that propose PTG homelands and dispersal trajectories on the basis of internal classifications of the TG family, as we do in this paper: Mello and Kneip (2017) and Rodrigues (2000).<sup>6</sup> While other works on the PTG homeland allude to linguistic facts, e.g., Lathrap's (1970:78) appeal to Arawakan loanwords in TG as evidence for a northern origin, or Urban's (1992) discussion of the geographic dispersal of Rodrigues' eight classic subgroups, they do not base their PTG homeland proposals on internal classifications of the family.

#### 4.1. Mello and Kneip (2017)

Mello and Kneip (2017:307) propose a PTG homeland subsuming the one we identify in this paper: a large ellipse spanning the lower Tapajós, Xingu, and Tocantins Rivers. They argue for their proposal using LMT, and the observation that this ellipse encompasses four of Rodrigues' (1984/5) eight classic TG *subconjuntos* (or five of nine subgroups, in Mello's (2000) modification of Rodrigues' classification),<sup>7</sup> making this area the locus of genealogical diversity of the family. Their conclusions are essentially as precise as the rake-like structure of Rodrigues' (1984/5) and Mello's (2000) classifications permit, and are broadly compatible with the proposal we advance in this paper, which specifies the lower Xingu as the PTG homeland.

Mello and Kneip (*ibid.*) propose three major dispersal trajectories: 1) a back-migration to the Rondônia area which corresponds to our Kayabí-Parintintin group; 2) a southwards coastal expansion by Tupinambá; and 3) a migration by the Guaranian and Bolivian TG languages (corresponding to our Southern group) southwards via Rondônia. These three proposals are broadly compatible with our own language dispersal, with the caveat that they are bolder than we are in proposing a specific migration route for the Southern languages. No clear evidence is presented in favor of this route over any other, however. It is also worth mentioning that since the Bolivian and Guaranian languages constitute three distinct subgroups in the classification that Mello and Kneip employ, they must in effect posit three independent migrations along this trajectory, an issue which they do not address. Note that in the Michael et al. (2015) classification, these three subgroups form a single subgroup, avoiding this difficulty.

In summary, Mello and Kneip's (2017) PTG homeland and dispersal trajectory proposals are broadly compatible with those proposed in this paper, with many of the differences being traceable to the fact that they base their application of LMT on a less finely-articulated internal classification of the family. Like us, they identify a northeastern homeland for PTG, based on the greater genealogical diversity of the family in that region,

---

<sup>6</sup> Schleicher (1998:320) proposes that the Planalto do Mato Grosso was the PTG homeland, not on the basis of an internal classification of the family *per se* (which he does not present), but on the spatial distribution of a number of phonological and morphosyntactic isoglosses (*ibid.*: 322). Although Schleicher (*ibid.*: 320) makes a nod to LMT, his conclusions do not result from LMT inferences based on a classification of the family.

<sup>7</sup> Mello and Kneip (2017:307) suggest that this area could even be considered to encompass six subgroups, depending on how far to the east the Tupinambá expansion may have begun from.

O'HAGAN, CHOUSOU-POLYDOURI, MICHAEL – PHYLOGENETIC...  
although the less fine-grained nature of the classification they employ does not facilitate their developing a more precise homeland proposal.

#### 4.2. Rodrigues (2000)

In contrast to Mello and Kneip (2017), Rodrigues (2000) proposes a southwestern Amazonian homeland for PTG that lies in the vicinity of the Arinos and upper Juruena River basins. Rodrigues alludes to the following evidence in support of his proposal: 1) that Rondônia, which lies relatively close to the west of this proposed homeland, is the region of greatest genealogical diversity of the Tupian family as a whole (including members of his TG *subconjunto* VI, that is, Michael et al's (2015) Kayabi-Parintintin group); and 2) certain phonological affinities between particular *subconjuntos* that Rodrigues suggests lend support to particular migratory scenarios. The latter phonological affinities are most cogently summarized in Rodrigues and Cabral (2002), which we discuss below.<sup>8</sup>

As we argue now, however, the evidence cited above does not in fact support the conclusion that the PTG homeland is located in Juruena-Arinos over alternative homeland hypotheses, including a northeastern homeland, as proposed in this paper.

First, Rodrigues' (2000) argument for a southwestern PTG homeland, to the degree that it incorporates LMT-based reasoning, is not framed in terms of the locus of genealogical diversity of the TG family, but rather, indirectly, on that of Tupian family as a whole. However, homeland inferences for the proto-language of a given group of languages should, according to LMT, be principally based on the locus of greatest genealogical diversity for that group of languages, and not that of languages higher up in the tree. Concretely, this means that the PTG homeland should be inferred principally on the basis of the locus of genealogical diversity of the TG subgroup, not on the basis of the locus of genealogical diversity of the larger Tupian family of which it is a part. As such, the fact that the proposed Arinos-Juruena homeland lies relatively near to the Proto-Tupian homeland is not, by itself, compelling support for this PTG homeland hypothesis.

Turning now to the issue of phonological affinities that Rodrigues (2000) mentions in support of the Arinos-Juruena PTG homeland proposal, it is useful to discuss Rodrigues and Cabral (2002), which updates Rodrigues' (1984/1985) classification, making modest changes to the membership of certain *subconjuntos* and, critically, adding higher-level structure to the classification of TG languages on the basis of morphological affinities and sound changes that they identify. Importantly, Rodrigues and Cabral (2002) address the same basic phonological affinities that Rodrigues (2000) presents to support the homeland and migration account he proposes, but they do so more explicitly, and in greater detail.

The higher-level structure that Rodrigues and Cabral (2002) propose is shown in Fig. 11, which compares this classification with that of Michael et al. (2015). Although

---

<sup>8</sup> Rodrigues (2000: 4-5) also discusses lexical borrowings from Cariban languages into TG ones, which Rodrigues argues must have occurred at the level of PTG, given their distribution of modern TG languages. Without a determination of where this language contact took place, however, these facts do not distinguish between different PTG homeland hypotheses. It should be noted that Mello and Kneip (2017:308) take evidence of TG-Cariban language contact as support for a northeastern PTG homeland, since that is closer to the ToC distribution of Cariban languages than the southwestern homeland proposed by Rodrigues (2000).

Rodrigues and Cabral (ibid.: 334-335) only make a passing allusion to the matter in this paper, they correctly observe that this classification is compatible with the southwestern origin and migration scenario proposed in Rodrigues (2000). This is due to the fact that: 1) members of two of the three major branches, encompassing three *subconjuntos*, are present in southwestern Amazonia, which, following LMT, can consequently be inferred to be the PTG homeland; and 2) all the other *subconjuntos*, most of which are found in northeastern Amazonia, form a single subgroup in this classification, allowing one to explain the modern distribution of these languages by positing a single migration from southwestern to northeastern Amazonia. What this demonstrates is that the higher-level structure of TG classifications is critical in distinguishing southwestern vs. northeastern Amazonian PTG homeland hypotheses.

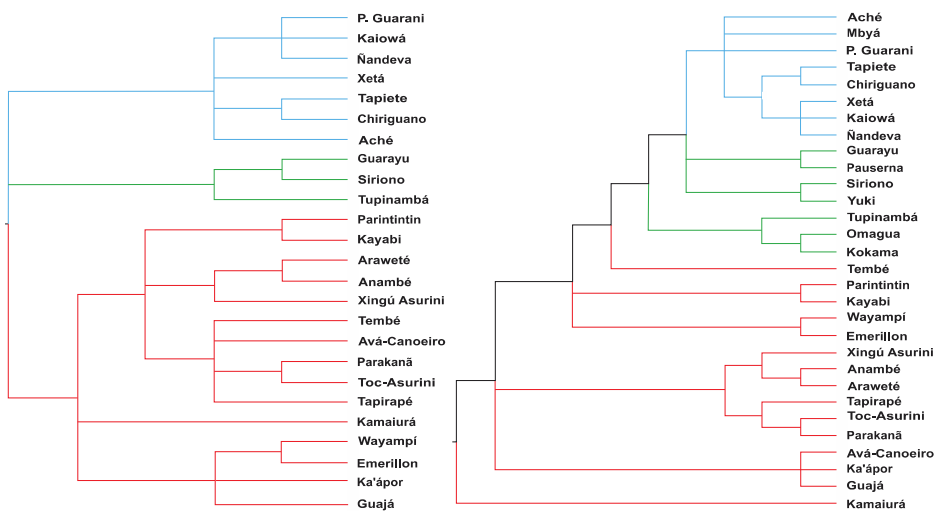


Fig. 11: Rodrigues and Cabral's (2002) classification of TG compared to Michael et al.'s (2015) classification

In this light, it is crucial to observe that Rodrigues and Cabral (2002) do not provide compelling evidence for the higher-level structure they propose, nor is it supported by the phylogenetic analysis, which does, however, support the traditional *subconjuntos*. In particular, the shared innovations that Rodrigues and Cabral present as evidence for the higher level structure in their proposed tree are that  $*ts > h(> \emptyset)$  in all *subconjuntos* but II and III, and that  $*tj > h(> \emptyset)$  in *subconjuntos* IV-VIII.<sup>9</sup>

Regarding this evidence, we first observe that none of these innovations group together *subconjuntos* II and III, meaning that no evidence is presented for this putative subgroup. In particular, note that the observation that languages in *subconjuntos* II and III did not experience lenition of  $*ts$  does not constitute the identification of a shared innovation, but rather, a shared retention, which is not evidence for subgrouping. And

<sup>9</sup> Rodrigues and Cabral also mention some morphological retentions, but since retentions do not provide evidence for subgrouping (which must be based on shared innovations), we do not consider them further.

critically, once we split the putative II+III subgroup into two distinct subgroups, there are then as many first order subgroups located *outside* of southwestern Amazonia as *within* it, already significantly weakening the basis for positing a southwestern Amazonian PTG homeland.

Second, we observe the claim that *\*ts>h(>ø)* did not affect *subconjuntos* II and III is rather misleading, aside from the fact that this would constitute a shared retention, rather than a shared innovation. This is because crosslinguistically, *\*ts* generally does not immediately debuccalize to *h* (i.e., *\*ts>h(>ø)*), but instead first lenites to *s*: *\*ts>s>h(>ø)*. Once we acknowledge that the sound change process in question is one in which lenition precedes debuccalization, we find that lenition also operated in languages of *subconjuntos* II and III. In particular, Sirionó of *subconjunto* II and Tupinambá of *subconjunto* III underwent lenition, but not Guarayú of *subconjunto* II, nor Kukama of *subconjunto* III. Even in terms of shared retentions, then, the languages of *subconjuntos* II and III do not pattern together, as Guarayú and Kukama are the only languages to retain *\*ts*. Moreover, as we can see, even languages in the same *subconjunto* can differ in terms of whether and when they underwent lenition.<sup>10</sup>

We are thus left with the question of the evidence supporting the large IV-VIII subgroup. As evident from the preceding paragraph, *\*ts>s* is not uniquely associated the IV-VIII subgroup, and even if it were, this lenition process is, again, so crosslinguistically common that it would not serve as compelling evidence for the subgroup. What the languages of IV-VIII do share is *s>h(>ø)*, but again, this is so common a sound change, that it has little probative value for subgrouping. Essentially similar arguments apply to the claim that *\*tj>h(>ø)* in all *subconjuntos* but I, II, and III.

In summary, then, Rodrigues and Cabral (2002) provide no compelling evidence for the higher-level structure they posit for the TG family, meaning that it cannot be adduced as evidence for a southwestern Amazonian homeland for PTG. Whatever the ultimate merits of Rodrigues' (2000) proposal for a southwestern Amazonian homeland, the linguistic evidence and argumentation provided for it is not compelling.

## 5. Comparing the Northeastern and Southwestern PTG homeland theories

Abductive reasoning, like that embodied by linguistic migration theory, is incapable of *proving* a proposition, instead yielding likely hypotheses that are intrinsically probabilistic in nature. For example, above we argued that the *lower* Xingu River basin was the most likely location of the PTG homeland, but it is certainly within the realm of possibility that it was located in the *middle* Xingu River basin. Alternative but very similar theories like these ultimately need to be evaluated by additional sources of evidence, such as the study of archaeological remains.

Despite the inherent probabilistic nature of abductive reasoning, it is typically feasible to evaluate the relative plausibility of starkly different alternative hypotheses, and it is this to which this section is dedicated. In particular, we compare the relative plausibility

---

<sup>10</sup> For example, in the Tupinambá-Omagua-Kukama subgroup, Kukama alone retains *\*ts*, indicating that Tupinambá and Omagua independently underwent *\*ts>s*, as might be expected with so common a sound change.



of the northeastern PTG homeland hypothesis (the ‘NE hypothesis’) that we defend in this paper to that of the southwestern homeland hypothesis (the ‘SW hypothesis’). As usefully summarized by Noelli (1998, 2008) in his overview of TG homeland proposals, one major tradition identifies the homeland as falling within Paraná River basin, which is the version of the SW hypothesis we compare against the NE hypothesis here. While others have developed alternative proposals as well (e.g., Almeida and Neves 2015), the SW hypothesis remains influential.

Before we begin, we stipulate an important constraint on the SW hypothesis we evaluate, with the goal of making it clearly distinct from the NE hypothesis we defend in this paper. Specifically, we require that the SW PTG homeland remain continuously occupied from the time at which PTG began to diversify to the modern era. The purpose of this restriction is twofold: first, this is consistent with the position taken by many defenders of the SW hypothesis, who cite early dates for ‘Guaraní’ remains in the Paraná River basin (see, e.g., Iriarte et al. (2017)), and second, it prevents the SW hypothesis from being trivially reduced to a hypothesis very similar to the NE hypothesis by positing an early migration of a high-level proto-language from the SW to NE regions. For example, imagine a version of the SW hypothesis that posits that while PTG was spoken in the SW region, PNTG migrated to the Xingu River basin after the first order split between NTG and Pre-Kamayurá. This latter hypothesis ends up being so similar to the NE hypothesis that it does not provide an insightful basis for comparison.

With this constraint, the following migration scenario is the most parsimonious one consistent with the SW hypothesis, Michael et al.’s (2015) classification, and the ToC distribution of TG languages:

- 1) PTG splits into PNTG and Pre-Kamayurá; Pre-Kamayurá migrates northeast to the upper Xingu River Basin.
- 2) NTG splits into Peripheral, Proto-Central and Proto-Tocantins. Proto-Central and Proto-Tocantins then independently migrate to the northeast, with Proto-Central settling and diversifying in the Xingu River Basin, and Proto-Tocantins doing so in the Tocantins River Basin.
- 3) Peripheral splits into Proto-Diasporic, Proto-Kayabí-Parintintin, and Proto-Emerillon-Wayampí. Proto-Kayabí-Parintintin migrates north, while Proto-Emerillon-Wayampí migrates northeast to the Xingu River basin, like Proto-Central did before it.
- 4) Diasporic then splits into Proto-Southern, Pre-Tembé, and POKT. Pre-Tembé and POKT then independently migrate to the northeast, each settling in the vicinity of the mouth of the Amazon, with POKT subsequently splitting into Pre-Tupinambá, which migrates back southwards along the coast, and Pre-POK migrating up the Amazon proper.
- 5) Finally, Proto-Southern diversifies, yield the ToC southern distributions of the languages of this large subgroup.

In light of Michael et al.’s (2015) classification, the migration scenario entailed by the SW hypothesis is considerably less plausible than that entailed by the NE hypothesis, as the former requires *six* independent migrations from the southwest to the northeast (Pre-Kamayurá, Proto-Central, Proto-Tocantins, Proto-Emerillon-Wayampí, Pre-Tembé, and POKT), including three independent migrations to the same river basin, i.e., Xingu



O'HAGAN, CHOUSOU-POLYDOURI, MICHAEL – PHYLOGENETIC...  
River basin (Pre-Kamayurá, Proto-Central, and Proto-Emerillon-Wayampí). While a hypothesis that requires independent but geographically correlated migrations of this sort is not intrinsically unbelievable,<sup>11</sup> it is, without strong additional evidence in its favor, considerably less plausible than one that does not require a set of independent but correlated migrations of this sort. In short, the NE hypothesis is considerably more plausible in light of Michael et al.'s (2015) classification than the SW hypothesis.

## 6. Conclusion

In this paper we have demonstrated that the application of Linguistic Migration Theory (LMT) to Michael et al.'s (2015) TG classification and the time of contact (ToC) distribution of TG languages yields the conclusion that the PTG homeland was located in the vicinity of the lower Xingu basin. We have also shown that the classification and distributions in question are not compatible with a southwestern homeland, which for purposes of explicitness, we took to be in the Paraná River basin, and which is the homeland for the Proto-Tupí-Guaraní people favored by many archaeologists on the basis of physical remains.

We do not, in this paper, seek to resolve the stark discrepancy between the homeland hypotheses favored by scholars working with different sources of evidence (i.e., linguistic and archaeological), but instead call attention to it, and identify it as a critical inter-disciplinary question to be addressed by scholars with overlapping interests regarding the deep social, cultural, and linguistic histories of the TG peoples. At the very least, these results call into question the assumption operative in much work on the topic (Noelli 1998: 649) that the distribution, diversification, and dispersal of TG languages mirrors that of the ceramics traditions associated with TG peoples. It is a truism that dates to early modern anthropology that culture, language, and populations have potentially distinct historical trajectories (Boas 1940; see also Donohue and Denham 2011), and it may be the case that we see evidence for significant differences among these trajectories in the case of TG peoples.

It is also worth noting, in this regard, that archaeological evidence has begun to accumulate that Tupí-Guaraní peoples have inhabited the lower Xingu and Tocantins basins for a considerable time (Almeida 2008; Garcia 2012), leading to proposals for an eastern, if not a northeastern, PTG homeland (Almeida and Neves 2015). Whether further archaeological work in the region will ultimately support an eastern or northeastern PTG homeland remains to be seen, but in light of the linguistics arguments presented in this paper, we suggest that such work should be considered a priority for TG archaeology.

By reversing the abductive processes leading to the hypothesis that the PTG homeland was located in the lower Xingu River basin, we have generated a set of hypotheses regarding the dispersal of the TG languages from the lower Xingu homeland. As described in §3, many of the higher-level splits in the diversification of the family are associated with relatively short-distance language dispersals, e.g., the split of PTG

---

<sup>11</sup> The migrations of different European peoples to the Americas could be an example of independent but correlated migrations of this sort. In this case, we explain the correlation as a result of the emergence of shared and competitive colonial practices among European nations.

into pre-Kamayurá and PNTG, with the former simply moving up the Xingu river, or the split of PNTG into Proto-Central, Proto-Tocantins, and Proto-Peripheral, which we argue was the result of the spread of PNTG to encompass the lower Xingu and lower Tocantins river basin, and the region east of the Tocantins River basin, with each of these second order daughters resulting from the separate development of PNTG in each of these major geographical areas.

Significantly, our analysis indicates that most of the considerable geographical dispersal of the TG languages is associated with the languages of the Peripheral subgroup, and especially the Diasporic subgroup of Peripheral. We argued that Proto-Peripheral diversified in the region east of the Tocantins River basin, with two of its daughters, Proto-Kayabi-Parintintin and Proto-Wayampí-Emerillon, moving eastward, and the third, Proto-Diasporic, continuing to diversify in the region east of the Tocantins River basin. Proto-Diasporic, in turn, split into three daughters, of which one, Pre-Tembé did not move significantly, but the other two, POKT and Proto-Southern, were associated with significant dispersals. POKT split into Pre-Tupinambá, which began a steady expansion southwards along the Atlantic coast, and Pre-POK, which moved far up the Amazon, subsequently experience significant language contact, which resulted in the emergence of POK proper (Michael 2014; O'Hagan 2011, 2019a,b). Proto-Southern moved an even greater distance to the Paraguay-Paraná River basin, diversifying and spreading there, resulting in the large Guaranian subgroup, centered on the Paraguay-Paraná River basin, and the smaller Siriono-Yuki and Warázu-Guarayu subgroups to the east, in the Guaporé and Mamoré River basins to the east. A major open question concerns the trajectory of the movement of Proto-Southern: the linguistic evidence at this point cannot distinguish between southwards migration along the coast, followed by an inland migration, or a migration along any of several major southern tributaries of the Amazon.

This account of the dispersal of TG languages constitutes a set of hypotheses, each of which both stimulates questions in affine disciplines such as archaeology and ethnohistory, and would benefit from evaluation by research in those disciplines. For example, as anticipated by Urban (1996), the fact that greatest geographical expansion of TG languages is localized in a particular subgroup raises the question of whether some social or cultural innovation at a particular node in the tree, say, Proto-Diasporic, drove or facilitated this expansion. Comparative ethnohistorical work and ethnographic work with the modern speakers of Diasporic languages may give us insight into this question. At the same time, our account provides a set of concrete hypotheses regarding the movements of languages, which, to the degree that they are associated with the movements of peoples and their material culture, can be evaluated by archaeological research.

Future work that may refine the PTG homeland and the dispersal account presented in this paper includes linguistic work that yields an even more finely-resolved TG tree, and the application of computational phylogeographical methods (see, e.g., Bouckaert et al. 2012), which would allow for probabilistic quantification of the homeland and dispersal trajectory hypotheses described in this paper.

## Abbreviations

ach: Aché; anb: Anambé; arw: Araweté; ava: Ava-Canoeiro; awe: Awetí; chr: Chiriguano; eme: Emérillon; Eme-Way: Emérillon-Wayampí subgroup; guaj: Guajá; gyu: Guarayú; kaa: Ka'apor; kai: Kaiowá; kam: Kamayurá; kay: Kayabí; kuk: Kukama; maw: Mawé; mby: Mbyá; NTG: Nuclear Tupí-Guaraní; ñan: Ñandeva; OKT: Omagua-Kukama-Tupinambá subgroup; omg: Omagua; PG: Paraguayan Guaraní; PTG: Proto-Tupí-Guaraní; PNTG: Proto-Nuclear Tupí-Guaraní; prk: Parakaná; prt: Parintintin; sir: Sirionó; tap: Tapirapé; TG: Tupí-Guaraní; tmb: Tembé; toc: Asuriní do Tocantins; ToC: time of contact; tpn: Tupinambá; tpt: Tapiete; way: Wayampí; war: Warázu; War-Gyu: Warázu-Guarayú subgroup; xet: Xetá; xing: Asuriní do Xingu; yuk: Yuki; Yuk-Sir: Yuki-Sirionó subgroup

## Acknowledgements

The authors would like to thank the audience of *Amazônicas V*, held in Belém, Brazil, 26-30 May, 2014, for helpful feedback on the proposals made in this paper.

## References

- Almeida, Fernando Ozorio de (2008). *O complexo tupi da Amazônia oriental* (Dissertação de mestrado). São Paulo: Universidade de São Paulo.
- Almeida, Fernando Ozorio de; Neves, Eduardo Góes (2015). Evidências arqueológicas para a origem dos tupi-guarani no leste da Amazônia. *Mana* 21(3): 499-525.
- Balée, William (1994). *Footprints of the forest: Ka'apor ethnobotany -- The historical ecology of plant utilization by an Amazonian people*. New York, NY: Columbia University Press.
- Blust, Robert (2013). *The Austronesian languages*. Canberra: Asia-Pacific Linguistics Research School of Pacific and Asian Studies, The Australian National University.
- Boas, Franz (1940). *Race, Language, and Culture*. Chicago: University of Chicago Press.
- Bouckaert, Remco; Lemey, Philippe; Dunn, Michael; Greenhill, Simon J.; Alekseyenko, Alexander V.; Drummond, Alexei J.; Gray, Russell D.; Suchard, Marc A.; Atkinson, Quentin D. (2012). Mapping the origins and expansion of the Indo-European language family. *Science* 337(6097): 957-960. doi: [10.1126/science.1219669](https://doi.org/10.1126/science.1219669)
- Brochado, José P. (1984). *An ecological model of the spread of pottery and agriculture into eastern South America* (Ph.D. dissertation). Champaign, IL & Urbana, IL: University of Illinois.
- Combès, Isabelle (2008). Los fugitivos escondidos: Acerca del enigma tapiete. *Bulletin de l'institut français d'études andines* 37(3): 511-533.
- Corrêa-da-Silva, Beatriz Carretta (2010). *Mawé/awetí/tupí-guaraní: Relações linguísticas e implicações históricas* (Tese de doutorado). Brasília: Universidade de Brasília.

- Diebold, Richard Jr. (1960). Determining the centers of dispersal of language groups. *International Journal of American Linguistics* 26(1): 1-10.
- Dyen, Isidore (1956). Language distribution and migration theory. *Language* 32(4): 611-626.
- Dimmendaal, Gerrit J. (2011). *Historical linguistics and the comparative study of African languages*. Amsterdam & Philadelphia: John Benjamins.
- Donohue, Mark; Denham, Tim (2011). Languages and genes attest different histories in Island Southeast Asia. *Oceanic Linguistics* 50(2): 536-542.
- Dumond, D.E. (1969). Toward a prehistory of the Na-Dene, with a general comment on population movements among nomadic hunters. *American Anthropologist* 71(5): 857-863.  
<https://doi.org/10.1525/aa.1969.71.5.02a00050>
- Dyen, Isidore (1956). Language distribution and migration theory. *Language* 32(4): 611-626.
- Galúcio, Ana Vilacy; Meira, Sérgio; Birchall, Joshua; Moore, Denny; Gabas, Nelson Jr.; Drude, Sebastian; Storto, Luciana; Picanço, Gessiane; Rodrigues, Carmen Reis (2015). Genealogical relations and lexical distances within the Tupian linguistic family. *Boletim do Museu Paraense Emílio Goeldi. Ciências Humanas* 10(2): 229-274.
- Garcia, Lorena Luana Wanessa Gomes (2012). *Arqueologia na região dos interflúvios Xingu-Tocantins: A ocupação tupi no Catefé* (Dissertação de mestrado). São Paulo: Universidade de São Paulo.  
doi: [10.11606/D.71.2012.tde-16052012-143440](https://doi.org/10.11606/D.71.2012.tde-16052012-143440)
- Grenand, Pierre (1982). *Ainsi parlaient nos ancêtres*. Paris: L'Office de la Recherche Scientifique et Technique Outre-Mer.
- Haynie, Hannah; Bower, Claire; Epps, Patience; Hill, Janel; McConnell, Patrick (2014). Wanderwörter in languages of the Americas and Australia. *Ampersand* 1: 1-18.  
<https://doi.org/10.1016/j.amper.2014.10.001>
- Heckenberger, Michael (2002). Rethinking the Arawakan diaspora: Hierarchy, regionality, and the Amazonian formative. In Jonathan D. Hill; Fernando Santos-Granero (eds.). *Comparative Arawakan histories: Rethinking language family and culture area in Amazonia*, pp. 99-122. Urbana, IL: University of Illinois Press.
- Horton, Donald (1948). The Mundurucu. In Julian H. Steward, *Handbook of South American Indians*, vol. 3, pp. 271-283. Washington, DC: United States Government Printing Office.
- Iriarte, José; Smith, Richard J.; Souza, Jonas Gregorio de; Mayle, Francis Edward; Whitney, Bronwen S.; Cárdenas, Macarena Lucia; Singarayer, Joy; Carson, John F.; Roy, Shovonlal; Valdes, Paul (2017). Out of Amazonia: Late-Holocene climate change and the Tupi-Guarani trans-continental expansion. *The Holocene* 27(7): 967-975. <https://doi.org/10.1177/0959683616678461>
- Lathrap, Donald W. (1970). *The Upper Amazon*. New York: Praeger.
- Meira, Sérgio; Drude, Sebastian (2015). A summary reconstruction of Proto-Maweti-Guarani segmental phonology. *Boletim do Museu Paraense Emílio Goeldi. Ciências Humanas* 10(2): 275-296.  
<http://dx.doi.org/10.1590/1981-81222015000200005>

## O'HAGAN, CHOUSOU-POLYDOURI, MICHAEL – PHYLOGENETIC...

- Mello, Antônio Augusto Souza (2000). *Estudo histórico da família linguística tupi-guarani: aspectos fonológicos e lexicais* (Tese de doutorado). Florianópolis: Universidade Federal de Santa Catarina. <http://repositorio.ufsc.br/xmlui/handle/123456789/78560>
- Mello, Antônio Augusto Souza; Kneip, Andreas (2017). Novas evidências linguísticas (e algumas arqueológicas) que apontam para a origem dos povos tupi-guarani no leste amazônico. *Literatura y Lingüística* 36: 299-312. <http://dx.doi.org/10.4067/S0716-58112017000200299>
- Michael, Lev (2014). On the pre-Columbian origin of Proto-Omagua-Kokama. *Journal of Language Contact* 7(2): 309-344.
- Michael, Lev; Chousou-Polydouri, Natalia (to appear). Computational Phylogenetics and the Classification of South American Languages. *Language and Linguistics Compass*.
- Michael, Lev; Chousou-Polydouri, Natalia; Bartolomei, Keith; Donnelly, Erin; Wauters, Vivian; Meira, Sérgio; O'Hagan, Zachary (2015). A Bayesian phylogenetic classification of Tupi-Guarani. *LIAMES- Línguas Indígenas Americanas* 15(2): 193-221. <https://doi.org/10.20396/liames.v15i2.8642301>
- Michael, Lev; O'Hagan, Zachary (2016). A linguistic analysis of Old Omagua ecclesiastical texts. *Cadernos de Etnolinguística* (Série Monografias 4).
- Nichols, Johanna (1997a). Modeling ancient population structures and movement in linguistics. *Annual Review of Anthropology* 26(1): 359-384.
- Nichols, Johanna (1997b). The epicenter of the Indo-European linguistic spread. In Roger Blench, Matthew Spriggs & John Terrell (eds.). *Archaeology and language I: Theoretical and methodological orientations*, pp. 122-148. London: Routledge.
- Nimuendajú, Curt (1924). Os índios parintintin do rio Madeira. *Journal de la Société des Américanistes* 16: 201-278.
- Nimuendajú, Curt (1948a). Tribes of the Lower and Middle Xingú River. In Julian H. Steward, *Handbook of South American Indians*, vol. 3, pp. 213-243. Washington, DC: United States Government Printing Office.
- Nimuendajú, Curt (1948b). The Maue and Arapium. In Julian H. Steward, *Handbook of South American Indians*, vol. 3, pp. 245-254. Washington, DC: United States Government Printing Office.
- Nimuendajú, Curt (1948c). The Cawahib, Parintintin, and their neighbors. In Julian H. Steward, *Handbook of South American Indians*, vol. 3, pp. 387-397. Washington, DC: United States Government Printing Office.
- Noelli, Francisco Silva (1998). The Tupi: Explaining origin and expansion in terms of archaeology and of historical linguistics. *Antiquity* 72(277): 648-663.
- Noelli, Francisco Silva (2008). The Tupi expansion. In Helaine Silverman; William H. Isbell, *Handbook of South American archaeology*, pp. 659-670. New York: Springer.
- O'Hagan, Zachary (2011). *Proto-Omagua-Kokama: Grammatical sketch and prehistory*. (BA thesis). Berkeley, CA: University of California, Berkeley.
- O'Hagan, Zachary (2019a). Restructuring of Proto-Omagua-Kukama kin terms. *Boletim do Museu Paraense Emílio Goeldi. Ciências Humanas* 14(1): 65-78.

- O'Hagan, Zachary (2019b). The origin of purpose clause markers in Proto-Omagua-Kukama. *Journal of Historical Linguistics* 9(2): 282-312.
- Pawley, Andrew; Ross, Malcolm (1993). Austronesian historical linguistics and culture history. *Annual Review of Anthropology* 22(1): 425-459.
- Ramirez, Henri; Vegini, Valdir; França, Maria Cristina Victorino de (2017). O warázu do Guaporé (tupi-guarani): Primeira descrição linguística. *LIAMES – Línguas Indígenas Americanas* 17(2): 411-506.  
<http://dx.doi.org/10.20396/liames.v17i0.8647468>
- Rodrigues, Aryon Dall'Igna (1984/1985). Relações internas na família linguística tupi-guaraní. *Revista de Antropologia* 27/28: 33-53.
- Rodrigues, Aryon Dall'Igna (2000). Hipótese sobre as migrações dos três subconjuntos meridionais da família tupi-guaraní. In *II Congresso Nacional da ABRALIN e XIV Instituto Linguístico*, 1596–1605. Florianópolis: Associação Brasileira de Linguística.
- Rodrigues, Aryon Dall'Igna; Cabral, Ana Suelly Arruda Câmara (2002). Revendo a classificação interna da família tupi-guaraní. In Ana Suelly Arruda Câmara Cabral; Aryon Dall'Igna Rodrigues (eds.). *Línguas indígenas brasileiras: fonologia, gramática e história*, pp. 327-337. Belém: Editora Universitária, Universidade Federal do Pará.
- Rodrigues, Aryon Dall'Igna; Cabral, Ana Suelly Arruda Câmara (2012). Tupian. In Lyle Campbell; Verónica Grondona (eds.). *The indigenous languages of South America: A comprehensive guide*, pp. 495-574. Berlin: de Gruyter Mouton.
- Rodrigues, Aryon Dall'Igna; Dietrich, Wolf (1997). On the linguistic relationship between Mawé and Tupi-Guaraní. *Diachronica* 14(2): 265-302.
- Santos-Granero, Fernando (2009). *Vital enemies: Slavery, predation, and the Amerindian political economy of life*. Austin, TX: University of Texas Press.
- Sapir, Edward (1916). *Time perspective in aboriginal American culture, a study in method*. Ottawa: Government Printing Office.
- Schleicher, Charles Owen (1998). *Comparative and internal reconstruction of Proto-Tupi-Guaraní* (Ph.D. dissertation). Madison, WI: University of Wisconsin, Madison.
- Urban, Greg (1992). A história da cultura brasileira segundo as línguas nativas. In Manuela Carneiro da Cunha (ed.). *História dos índios no Brasil*, pp. 87–102. São Paulo: Fundação de Amparo à Pesquisa do Estado de São Paulo, Companhia das Letras & Secretaria Municipal de Cultura.
- Urban, Greg (1996). On the geographical origins and dispersions of Tupian languages. *Revista de Antropologia* 39(2): 61-104.

Recebido: 25/6/2019

Versão revista e corrigida: 23/10/2019

Aceito: 29/10/2019